Synaptic theory of gradient learning with empiric inputs

Ila Fiete

Kavli Institute for Theoretical Physics

Types of learning

- Associations among events (and reinforcements).
  e.g., classical conditioning
- How one's actions affect events (and reinforcements), how to shape actions.
  e.g., instrumental conditioning

behavioral	neural
classical conditioning	Hebbian learning
instrumental conditioning	?

Outline

- Learning with reinforcements as optimization
- Synaptic learning rule
- Scale-up problem
- Example: birdsong
- Implications

with Sebastian Seung

Instrumental conditioning

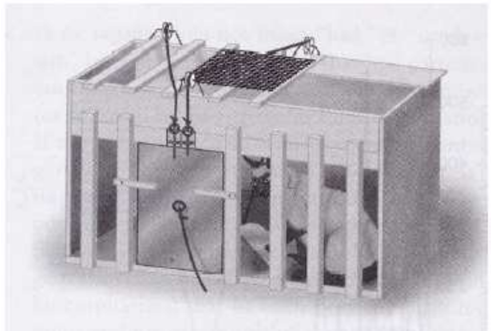
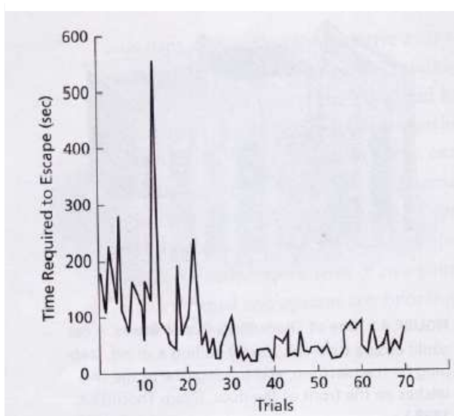


FIGURE 4.1 One of Thorndike's Puzzle Boxes. A cat could escape from this box by pulling a string, stepping on the platform, and turning one of the two latches on the front of the door. (From Thorndike, 1898.)

Learning by reinforcement



Thorndike, 1898

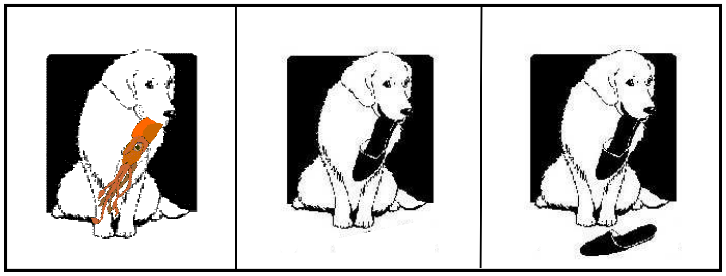
Try multiple strategies; modify behavior in way that tends to *improve* reinforcement.

The law of effect

Of several responses made to the same situation those which are accompanied or closely followed by *satisfaction* to the animal will become *more likely to recur*; those which are accompanied or closely followed by *discomfort* to the animal will, other things being equal, have their connections to the situation weakened, so that, when it recurs, they will be less likely to *occur*. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.

Thorndike, 1910

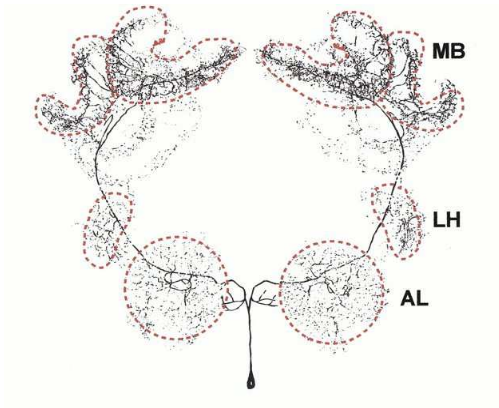
Trial, reward, and optimization



reward = performance on task

behavioral	neural
classical conditioning	Hebbian learning
instrumental conditioning	?

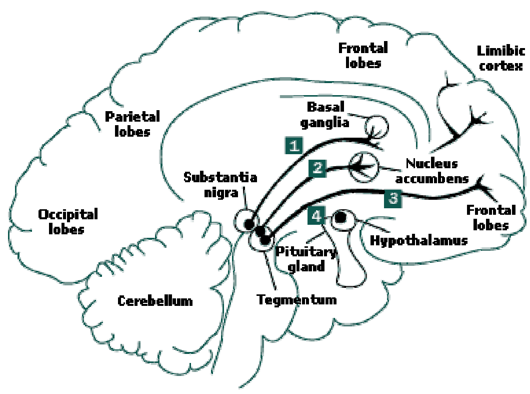
Octopamine system



Insect olfactory learning

VUMmx1

Dopamine system



Neural basis of learning

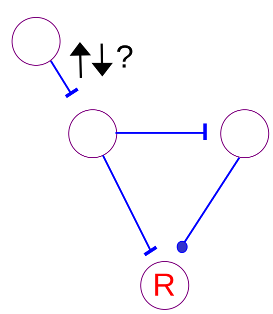
Global signal
(Reward, motor error, etc.)

Local signals
(Voltage, calcium, etc.)

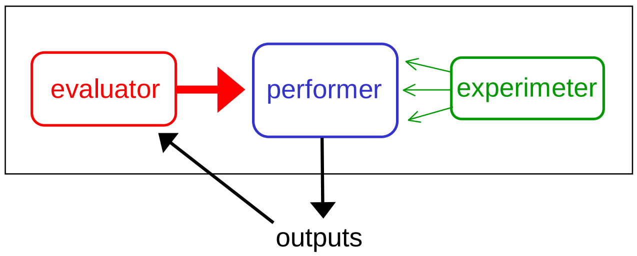
Synaptic plasticity

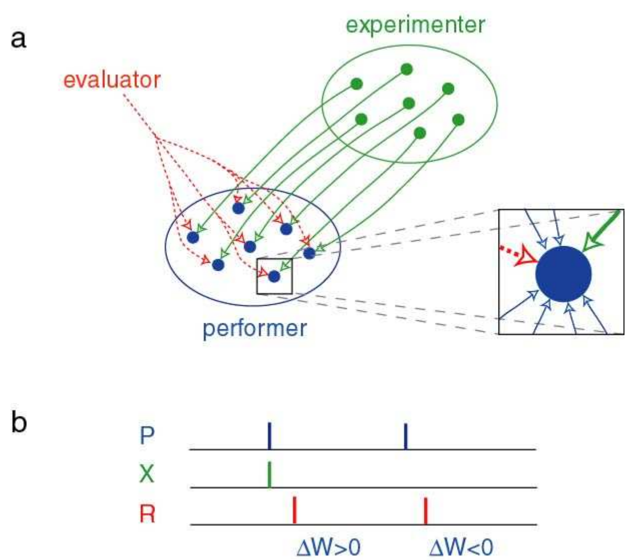
The interaction between global and local signals is largely uncharacterized.

Why difficult?



Learning with empiric inputs



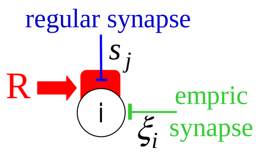


Fiete and Seung

learning rule

$$\Delta W_{ij} = \eta R e_{ij}$$

$R = R[V]$



$$g_i^{syn} = \sum_j W_{ij} s_{ij} + \xi_i$$

$$e_{ij} = \int_0^T dt \left(\xi_i - \left\langle \xi_i \right\rangle\right) s_{ij}$$

local eligibility

gradient following

$$\left\langle \Delta W_{ij} \right\rangle = \frac{\partial \left\langle R \right\rangle}{\partial W_{ij}}$$

example: spiking network

$$C_m \frac{dV_i}{dt} = -g_i^{syn}(V_i - V_E) - \sum_\alpha g_{\alpha i}^{ionic}(V_i - V_\alpha)$$

$$\frac{ds_{ij}}{dt} = -f_s(V, s, \{u_{se}, \ldots\})$$

$$g_i^{syn} = \sum_j W_{ij} s_{ij} + \xi_i$$

$$\Delta W_{ij} \stackrel{?}{=} \left\langle R s_{ij} \right\rangle$$

Sensitivity lemma

$$\frac{\partial \left\langle R \right\rangle}{\partial W_{ij}} = \frac{\partial}{\partial b_i} \left\langle R s_{ij} \right\rangle$$

$$\left(g_i^{syn} = \sum_j W_{ij} s_{ij} + b_i + \xi_i \right)$$

Stochastic gradient ascent

-Reward optimization, but only on average.
-Model-free.
-Local to synapses aside from evaluation.
-Deals with delayed evaluation and dynamic synapses.
-Slow?

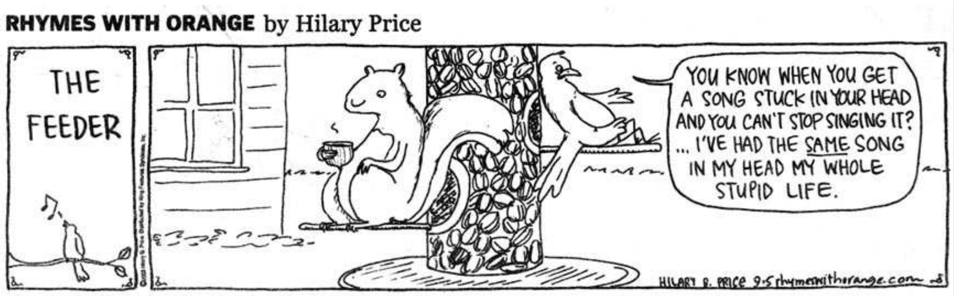
Learning time

- Individual correlations small → long averaging?

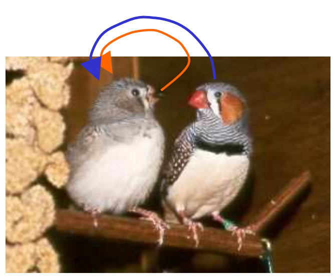
- Apply to biological example.

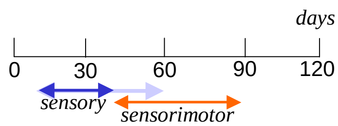
Birdsong learning

with Sebastian Seung & Michale Fee

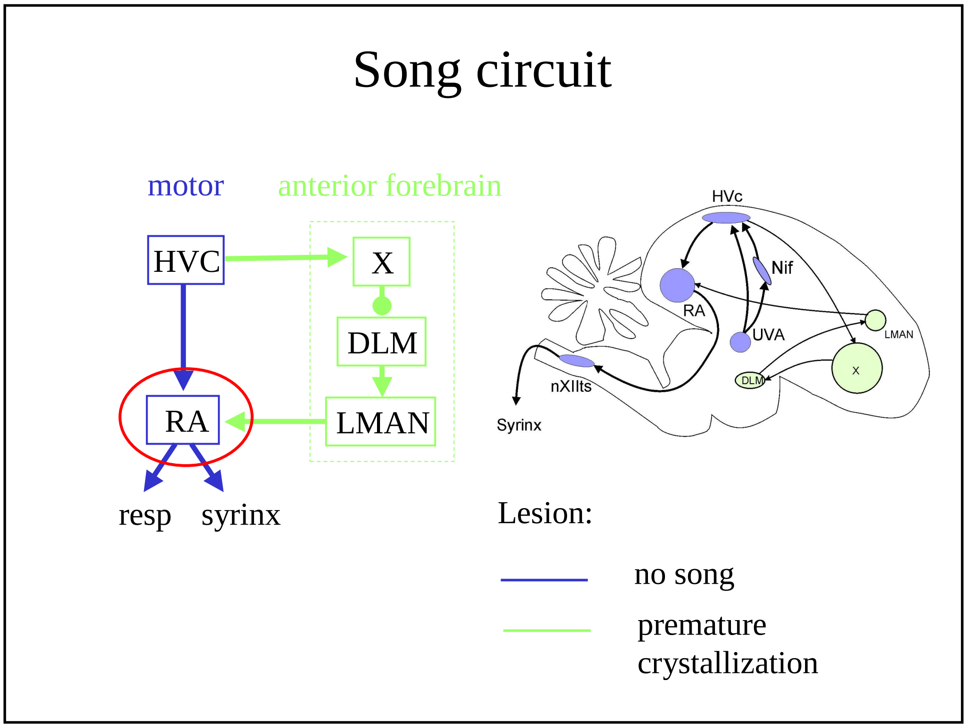


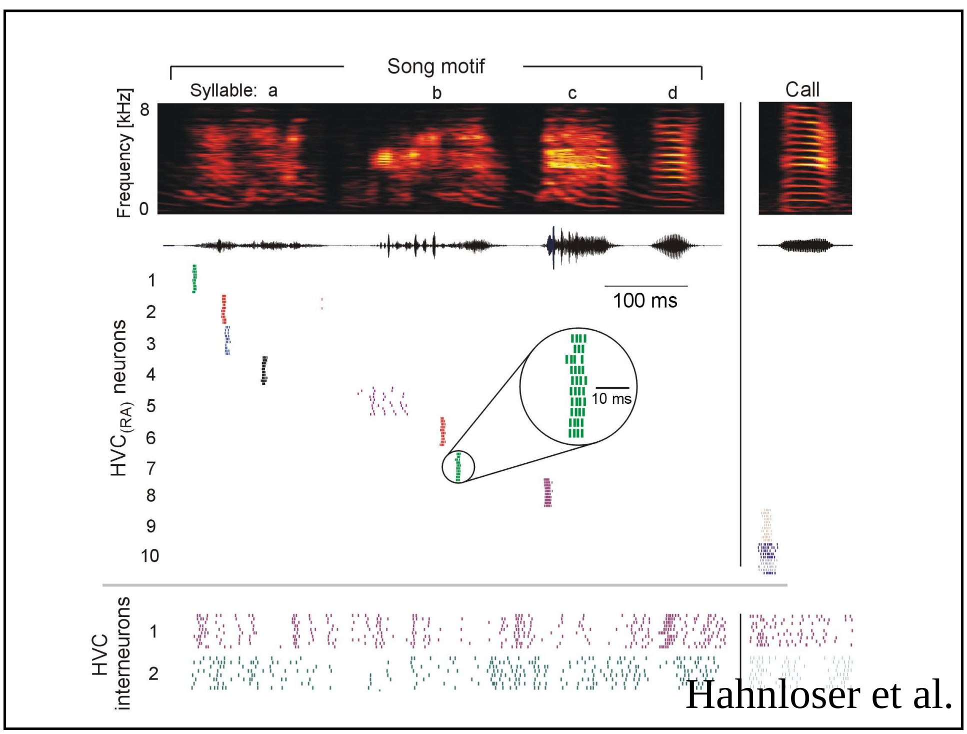
Behavior



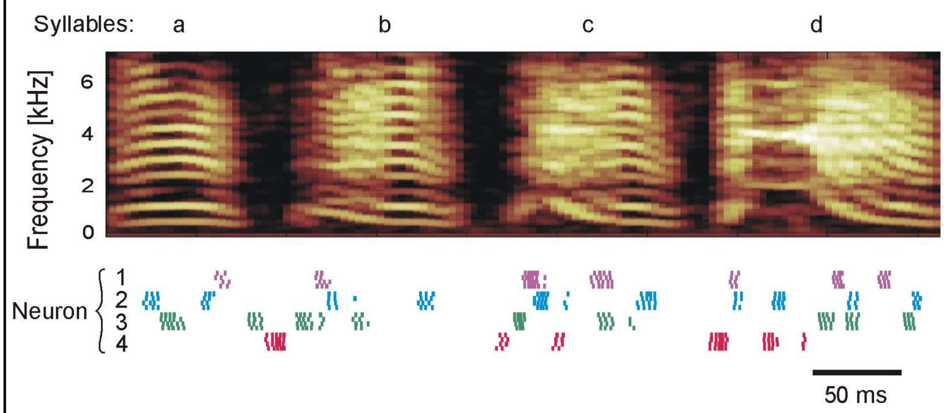


Social feedback and tutor song not needed.
Auditory feedback of own song crucial.

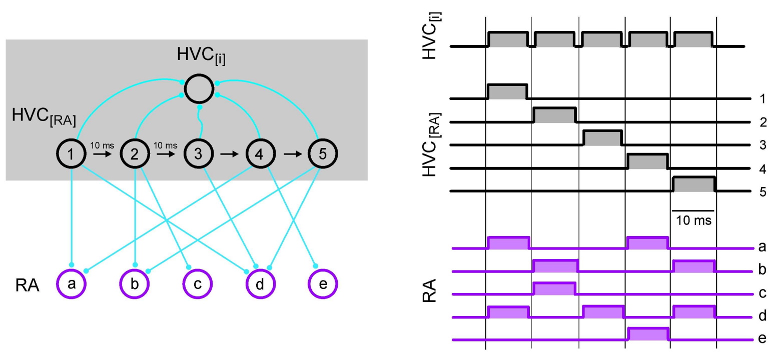

Song circuit
motor
anterior forebrain
HVC
X
DLM
RA
LMAN
resp
syrinx
HVc
Nif
RA
UVA
LMAN
X
DLM
nXIIts
Syrinx
Lesion:
no song
premature
crystallization


Song motif
Syllable: a
b
c
d
Call
Frequency [kHz]
8
0
100 ms
10 ms
HVC(RA) neurons
1
2
3
4
5
6
7
8
9
10
HVC interneurons
1
2
Hahnloser et al.

RA activity

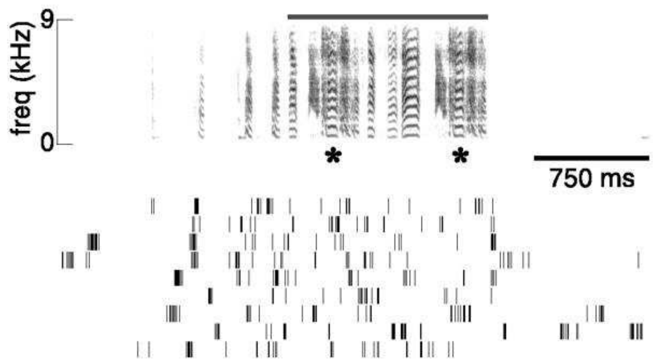


Song representation



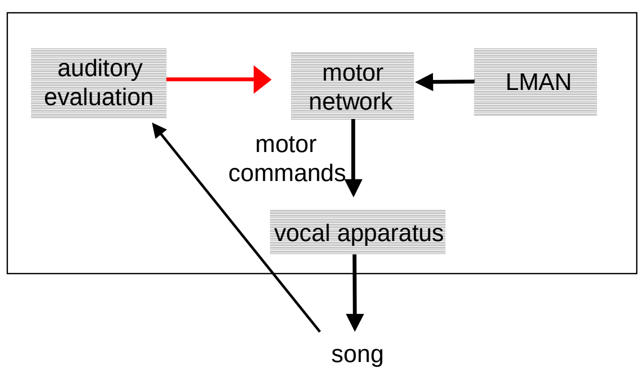
HVC generates a sparse sequence
later stages learn/perform the motor map

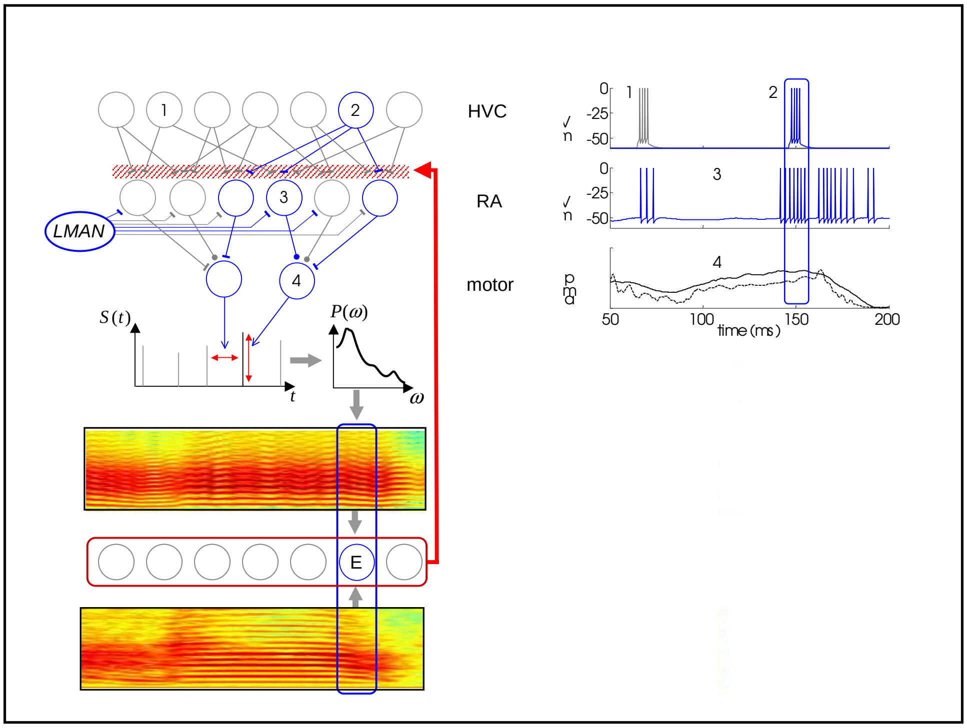
LMAN activity is variable

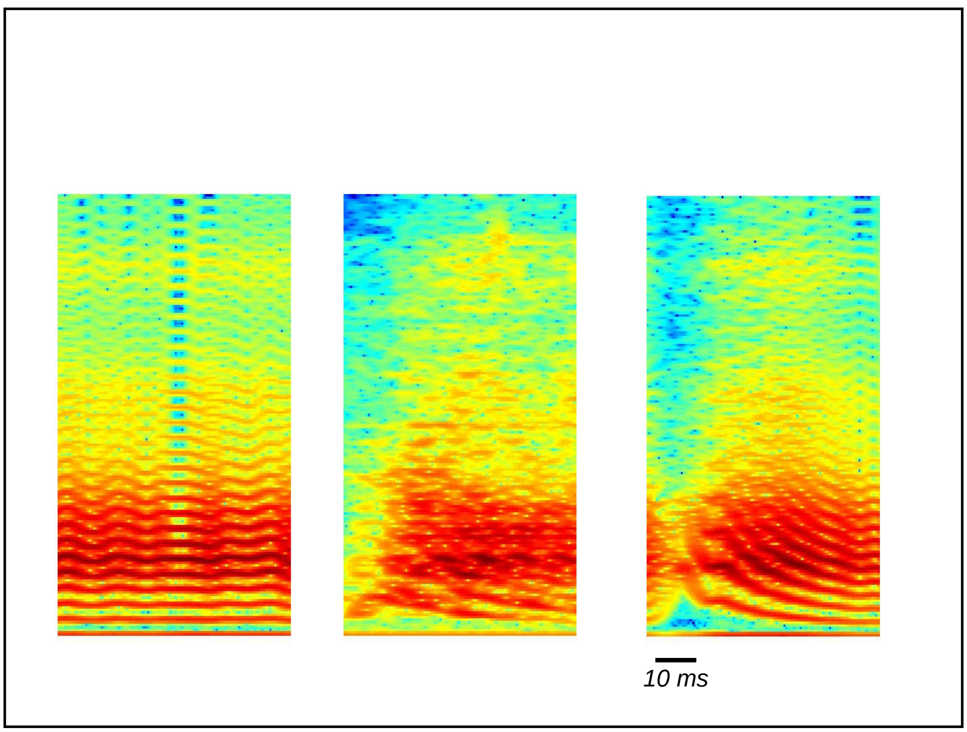


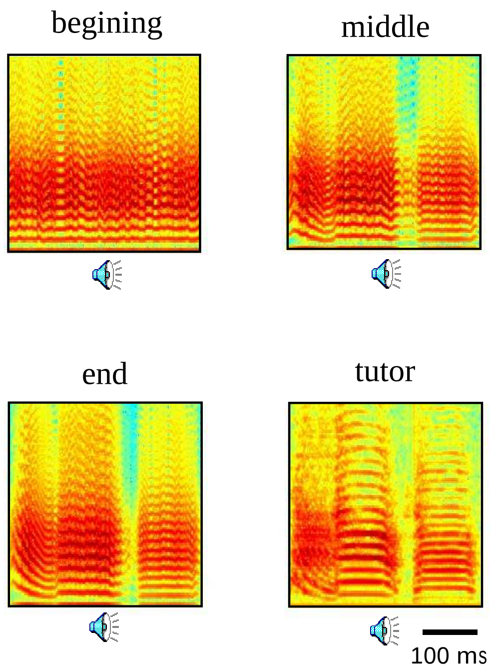
Hessler & Doupe

Song system schematic




HVC
RA
motor
LMAN
1
2
3
4
E
$S(t)$
t
$P(\omega)$
ω
0
-25
-50
50
100
150
200
time (ms)


10 ms



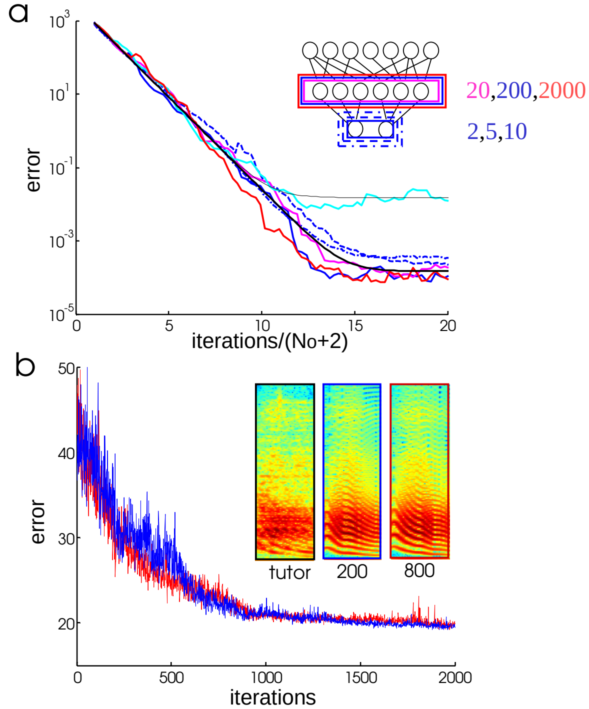
2000 epochs

Scaling to large networks

Correlation of single neuron with reward = 1/N.
Time to learn ~ N?

factor of 4000 more neurons in bird:

	HVC	RA	outputs
model	800	200	2
bird	20000	8000	8



Redundancy and rank

Correlation of single output with reward = $1/N_o$.

Time to learn ~ N_o.

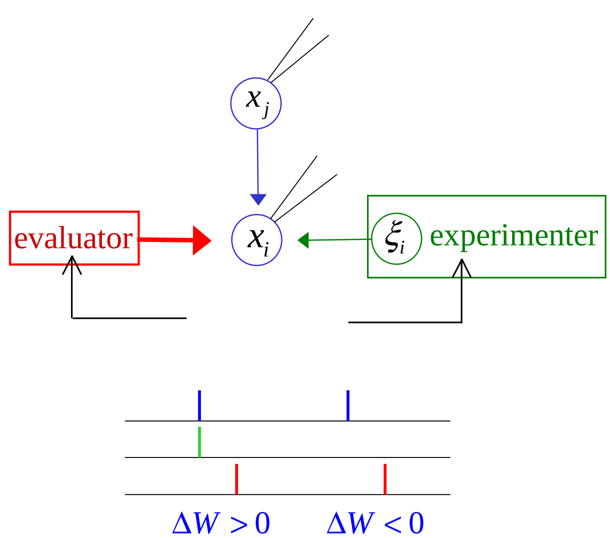
Model: 2 outputs
Bird: 8 outputs

Neural or synaptic noise. Time: <8000 epochs.

‘Simple’ tasks can be learned

- Learning time is independent of network size in redundant networks.
- Zebra finches practice 80,000 times.
- Motor control is redundant and learning is often slow.

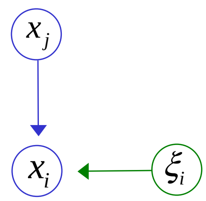
Experimental probe



Why is the brain so noisy?

- Why is spiking irregular?
- Why are synapses unreliable?
- Does the brain use noise for learning?

Similarity to Hebbian learning?



$$\Delta W_{ij} = R\left(\xi_i - \left\langle \xi_i \right\rangle\right) s_j$$

strong noise input and weak regular inputs

$$\approx R\left(\delta_i - \left\langle \xi_i \right\rangle\right) s_j$$

need segregated noise inputs so noise baseline correct

Supervised, reinforced, or unsupervised?

noise as exploration

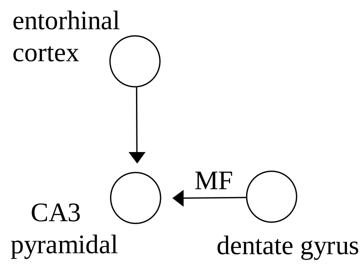
↓

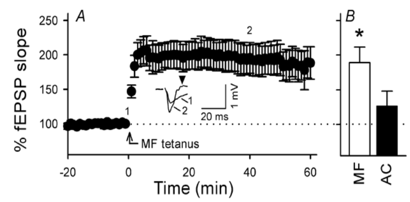
noise as instruction

post noise spike $\xrightarrow{\text{reward}}$ increase active weights
no noise spike $\xrightarrow{\text{reward}}$ decrease active weights

however, ‘instruction’ is stochastic, uncorrelated with error

Evidence of heterosynaptic plasticity

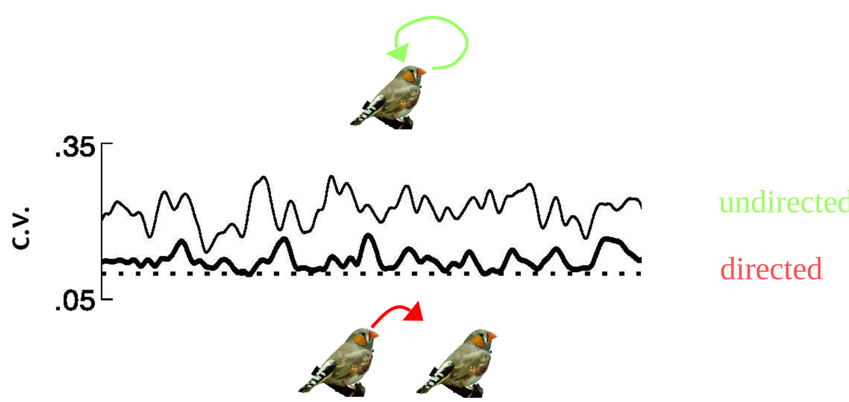




Tsukamoto et al., 2003

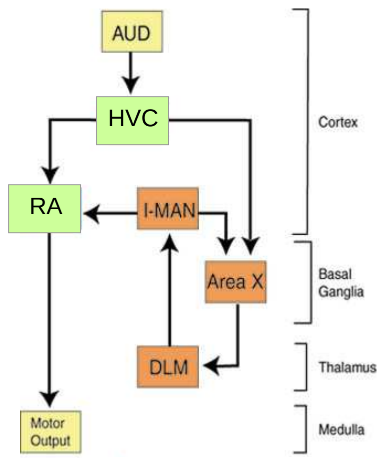
also cerebellar CF, PF convergence at Purkinje cells

Practice vs. performance

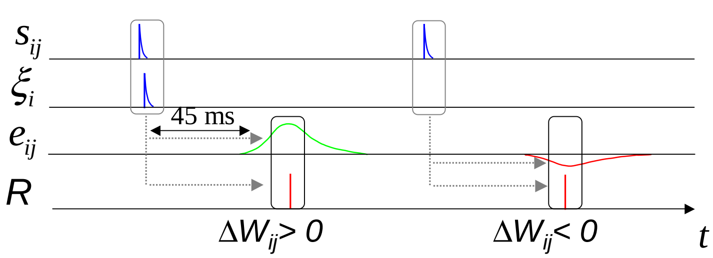


Hessler and Doupe, 1999

Song circuit



Delayed reward

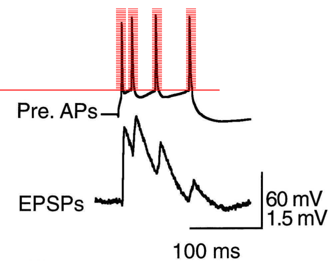


IF dynamics

$$C_m \frac{dV_j}{dt} = -g_L(V_j - V_L) - g_{Ej}(V_j - V_E)$$

$$V_j = V_{thres}$$

$$V_j := V_{reset}$$
$$s_j := s_j + \Delta$$



$$\tau_s \frac{ds_j}{dt} = -s_j(t)$$

$$g_{Ei} = \sum_j W_{ij} s_j(t)$$

Gradient following:

$$\Delta R = \frac{\partial R}{\partial W_{ij}} \Delta W_{ij}$$

$$\Delta W_{ij} = \frac{\partial R}{\partial W_{ij}} \Rightarrow \Delta R = \left(\frac{\partial R}{\partial W_{ij}} \right)^2 > 0$$

REINFORCE:

$$\langle \Delta W_{ij} \rangle = \frac{\partial \langle R \rangle}{\partial W_{ij}} = \frac{\partial}{\partial W_{ij}} \sum_s P(s) R(s)$$

$$= \sum_s P(s) \left(R(s) \frac{\partial \ln P(s)}{\partial W_{ij}} \right)$$

$$= \left\langle R(s) \frac{\partial \ln P(s)}{\partial W_{ij}} \right\rangle$$

$$\Delta W_{ij} = R(s) \frac{\partial \ln P(s)}{\partial W_{ij}} \equiv R e_{ij}$$